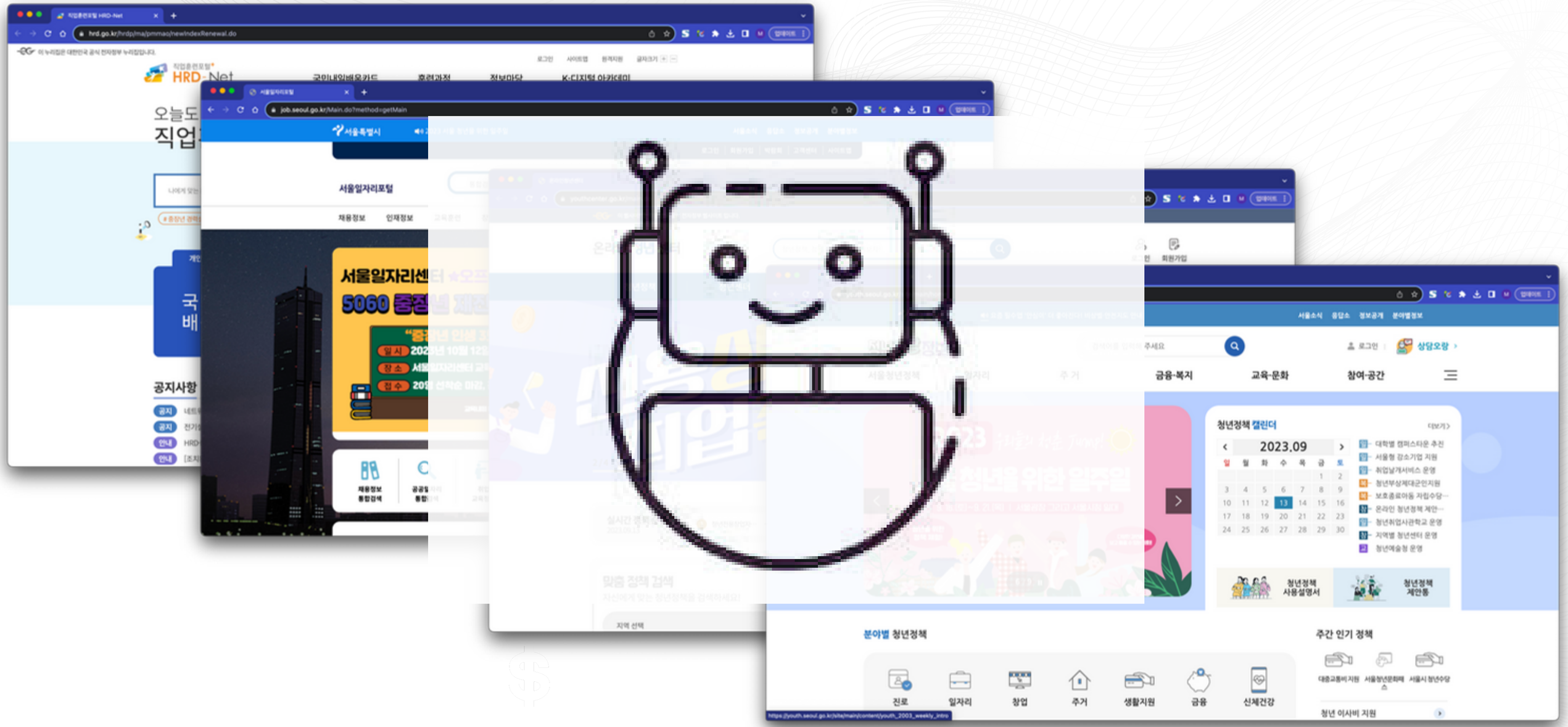
# CHAT-PUB

Helper of welfare policy for youth

김강산
양승빈
박진호
전창민

# TABLE OF CONTENTS

# INTRODUCTION



# INTRODUCTION

## Our visions...

1

### Primary Goal ☑

- A policy recommendation service that allows information on youth policies
- Create a highly accurate generative Chatbot using the NLP model
- Targeting young people living in Seoul and Gyeonggi Province

2

### Secondary Goal

- A rating system community
- Develop dedicated website
- Expanding service coverage to the metropolitan areas

3

### Last Goal

- A more detailed guide to the policy you want to proceed with
- Introduction to additional information required during the application process
- Nationwide service

# PROJECT PROGRESS

| Task | week 3 | week 4 | week 5 | week 6 | week 7 | week 8 | week 9 | week 10 | week 11 | week 12 | week 13 | week 14 | week 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Topic Study | | | | | | | | | | | | | |
| Build Data Pipeline | | | | | | | | | | | | | |
| Modeling | | | | | | | | | | | | | |
| Front-End | | | | | | | | | | | | | |
| Back-End | | | | | | | | | | | | | |
| Testing | | | | | | | | | | | | | |

# TEAM MEMBERS

## Kangsan Kim, Team Leader, UI/UX

- Designing better User Experience
- Visual Design
- Developing Web Page
- Figma, React

## Jinho Park, Data Engineer

- Data Pipeline
  - Crawling/Preprocessing/Generation ...
- Database Integration
- Code Refactoring
- Python, DBeaver

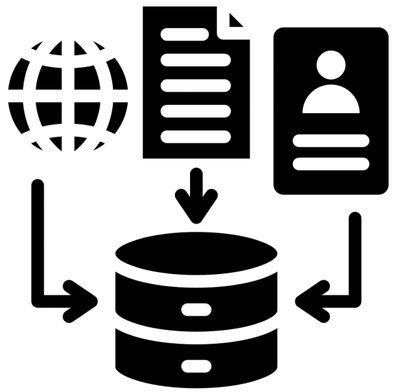
## Changmin Jun,  Back-End Developer

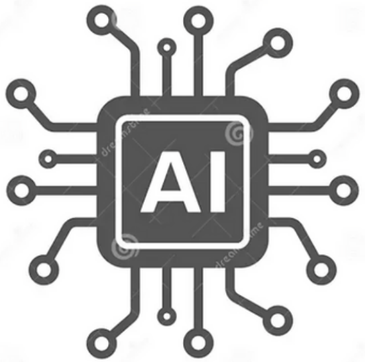
- Construct API
- Desgin and Connect DB
- Developing server using EC2, NGINX
- React, MariaDB, Fast API

## Seungbin Yang, AI Engineer

- Model Development
- Model Training & Evaluation
- Paper Search
- Python, Pytorch

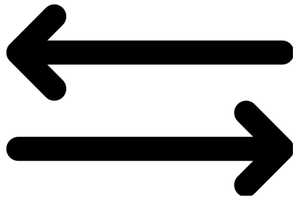
# IMPLEMENTATION






- Data
- Model
- Back-End
- Front-End

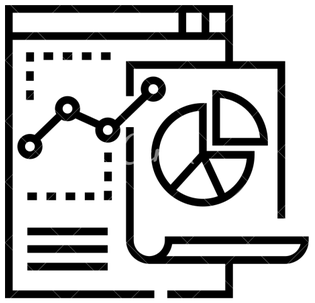
‹Main Functionality›



‹Web Services›

# DATA

## DATA PIPELINE



**Raw DATA**

EDA(Exploratory Data Analysis)

Implement code for Data Crawling

Data Preprocessing

Data Integration with Database

Automation & Monitoring

**Refined DATA**

# DATA PIPELINE OF CHATPUB

# DATA

## DATA PIPELINE

There are some challenges...

## Mainly Focus on...

1. How to refine RAW DATA?
2. How to make Training Datasets?

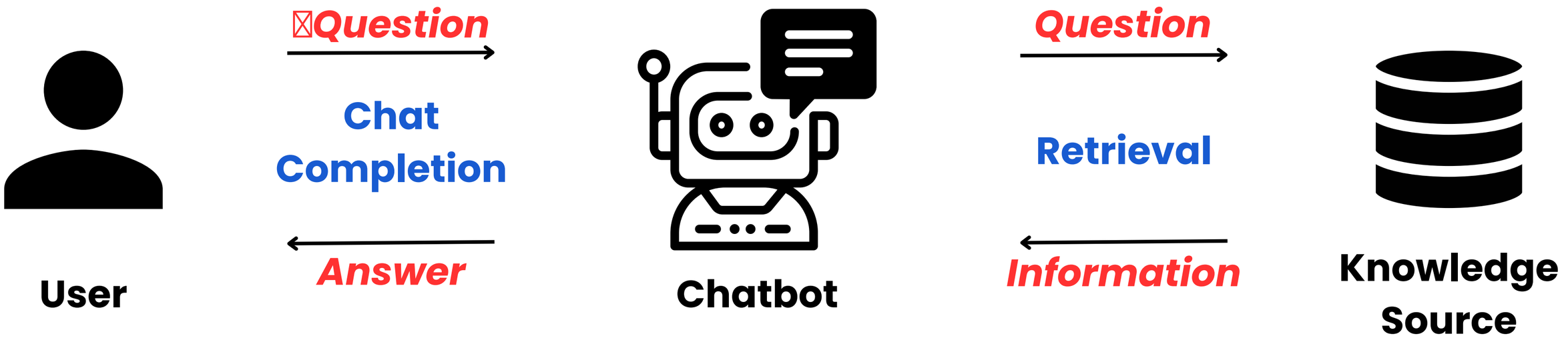
-> USE FAISS !!


```
search(model, index, '구리시에 관심이 있는데 괜찮은 정책 있어?', 1)
✓ 0.1s

[[ 2.22022831e-02 -5.36617100e-01 -8.72454569e-02 -6.93040669e-01
   1.95074722e-01 -1.01507910e-01  8.59122872e-01 -1.15090422e-01
   3.60790044e-01 -4.36832309e-02 -2.85417706e-01 -2.49668226e-01
   3.11255962e-01  2.12615371e-01 -1.44003153e-01 -4.72312391e-01
   2.29062319e-01 -7.68211007e-01  2.83244610e-01 -1.10166359e+00
  -1.32374123e-01  6.76300451e-02 -2.11506173e-01  3.73165399e-01
  -1.00836039e+00  6.11341655e-01  6.28062263e-02  5.42643726e-01
  -6.65344298e-01  4.01938856e-01  3.73282760e-01 -3.87211710e-01
  -1.28040284e-01 -6.10775352e-01 -1.07950374e-01  6.16771996e-01
  -3.65155518e-01 -6.71888828e-01  3.09403148e-02  2.73636281e-02
  -6.52105689e-01 -3.30289453e-01 -2.31157631e-01  3.83417867e-02
   4.56821054e-01 -8.11878219e-02 -1.60433561e-01 -5.22067487e-01
  -4.73977953e-01  3.92300069e-01  2.55464017e-01 -7.35937208e-02
   1.92174111e-02  5.84220409e-01  2.93138891e-01 -3.57280940e-01
   4.07842398e-01 -1.53997496e-01 -4.41726208e-01  2.25594983e-01
   4.25655574e-01  2.78181404e-01  6.44712523e-02  2.45962977e-01
   5.09047834e-03 -3.41961235e-01 -8.38873804e-01 -1.16708264e-01
  -8.86502802e-01  2.51761079e-02 -7.09160447e-01 -1.90470845e-01
   9.20764089e-01  4.97447550e-01  2.15500265e-01 -6.28849030e-01
  -1.88646719e-01  6.23598456e-01  1.08124125e+00 -8.47105756e-02
  -5.74157357e-01 -5.36367074e-02 -2.01641440e-01 -1.99187666e-01
  -4.24037836e-02  8.57522130e-01 -2.33404730e-02  2.58832037e-01
   2.82135308e-01 -4.09846485e-01  8.96249190e-02  4.12386889e-03
  -1.73793897e-01 -5.38320206e-02 -5.77122569e-01  1.66340634e-01
  -1.62463889e-01 -1.01045065e-01  1.23788230e-01  1.73807904e-01
...
summary: {'정책 번호': 'R2023060212999', '정책 분야': '교육분야', '지원 내용':
qualification: {'연령': '만 18세 ~ 39세', '거주지 및 소득': '사업개시일 현재 만18
methods: {'신청 절차': '구비 서류 양식 작성 및 증빙서류 구비', '심사 및 발표': 'ㅇ
```

# MODEL

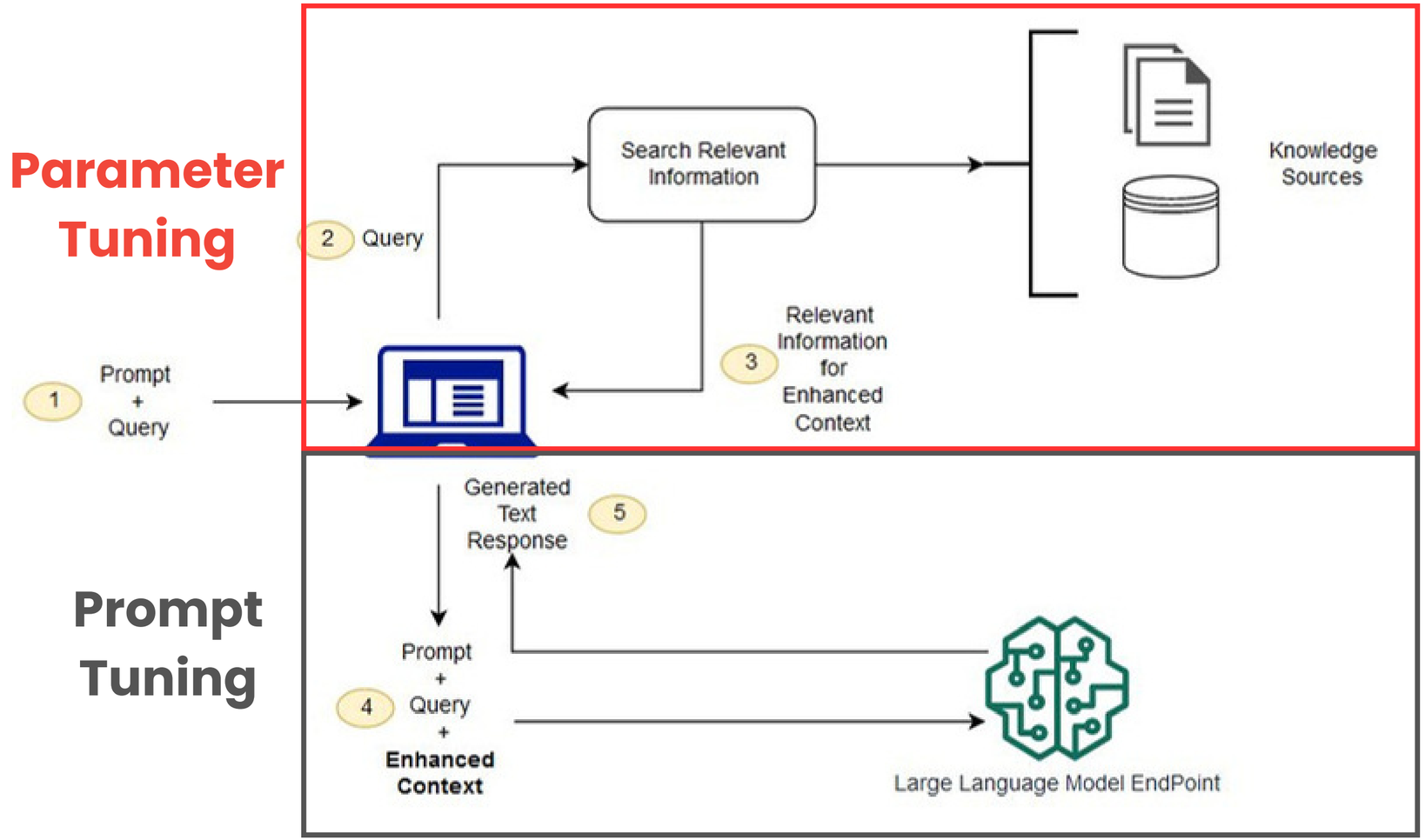
## RECAP: RAG




**Retrieval Augmented Generation**

# MODEL

## RECAP: RAG




[Retrieval]

Sentence Transformers

[Generation]

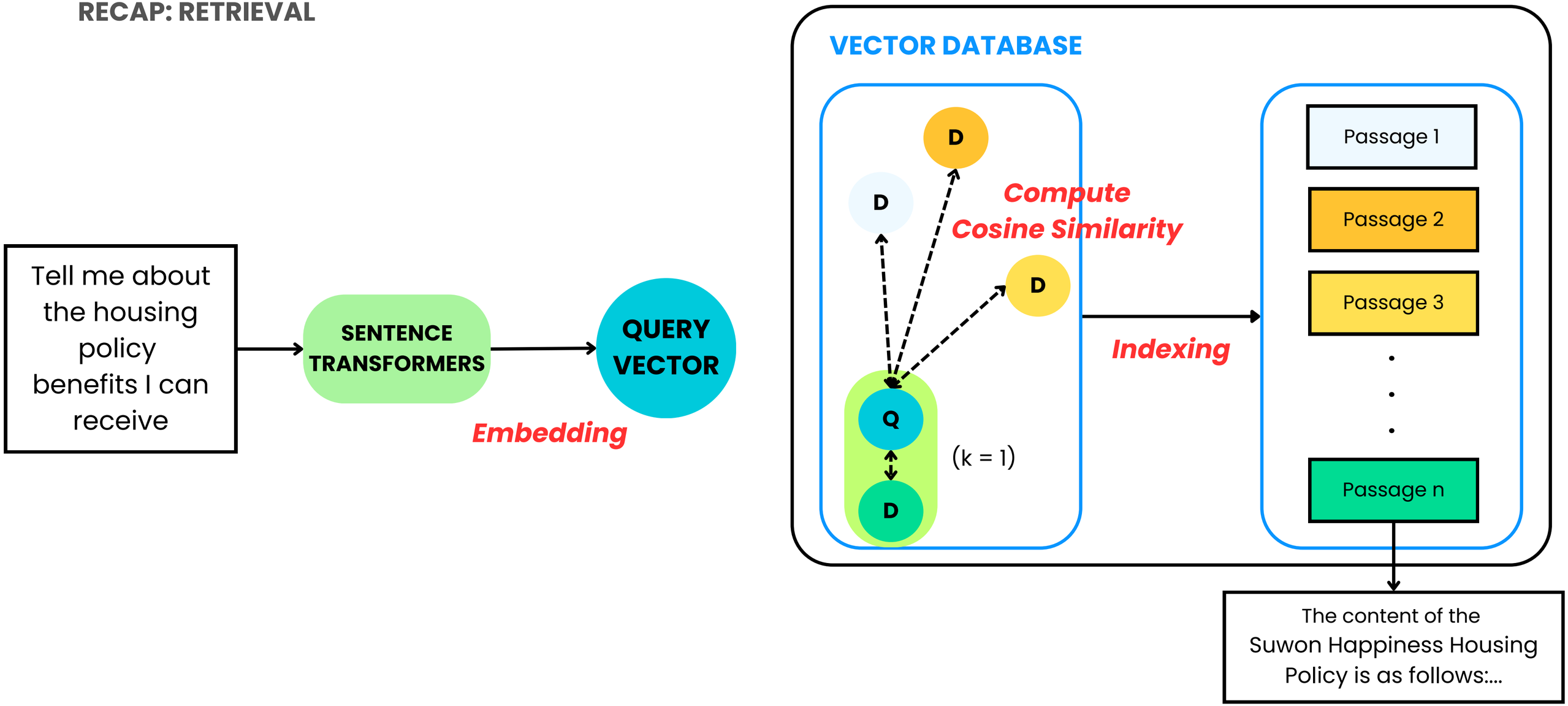
Open AI API (GPT-4-turbo)

HTTPS://DOCS.AWS.AMAZON.COM/SAGEMAKER/LATEST/DG/JUMPSTART-FOUNDATION-MODELS-CUSTOMIZE-RAG.HTML

# MODEL

## RECAP: RETRIEVAL

KNOWLEDGE SOURCES

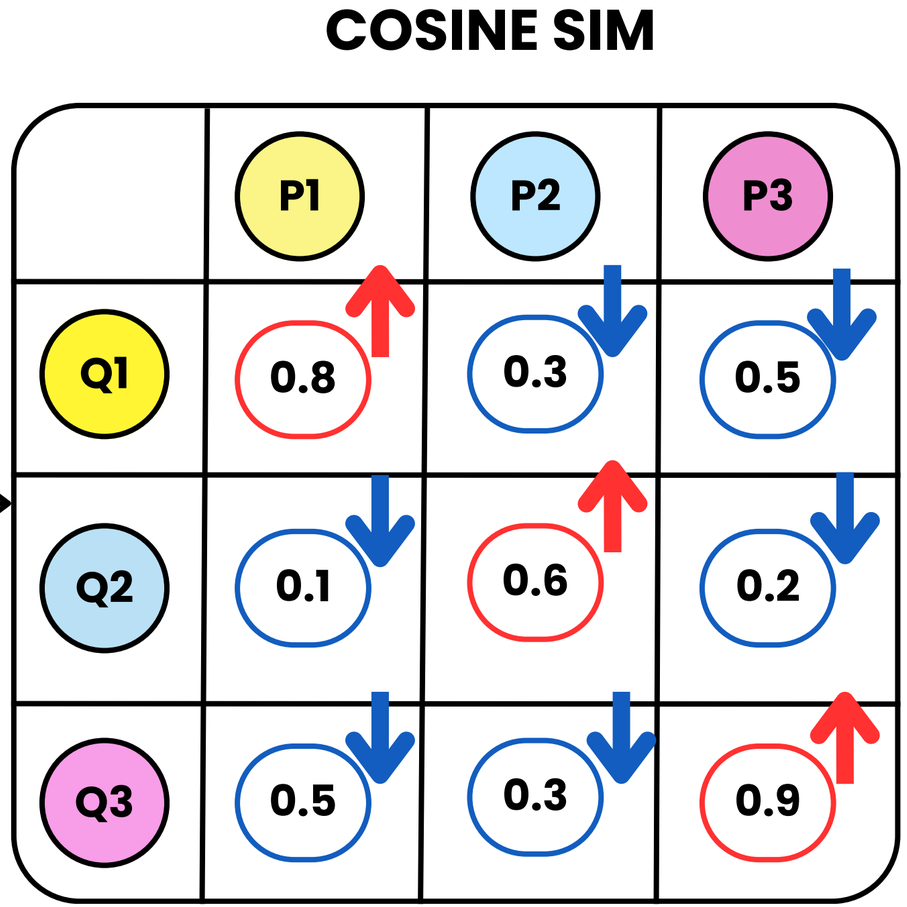


# MODEL

## RECAP: SENTENCE TRANSFORMERS

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

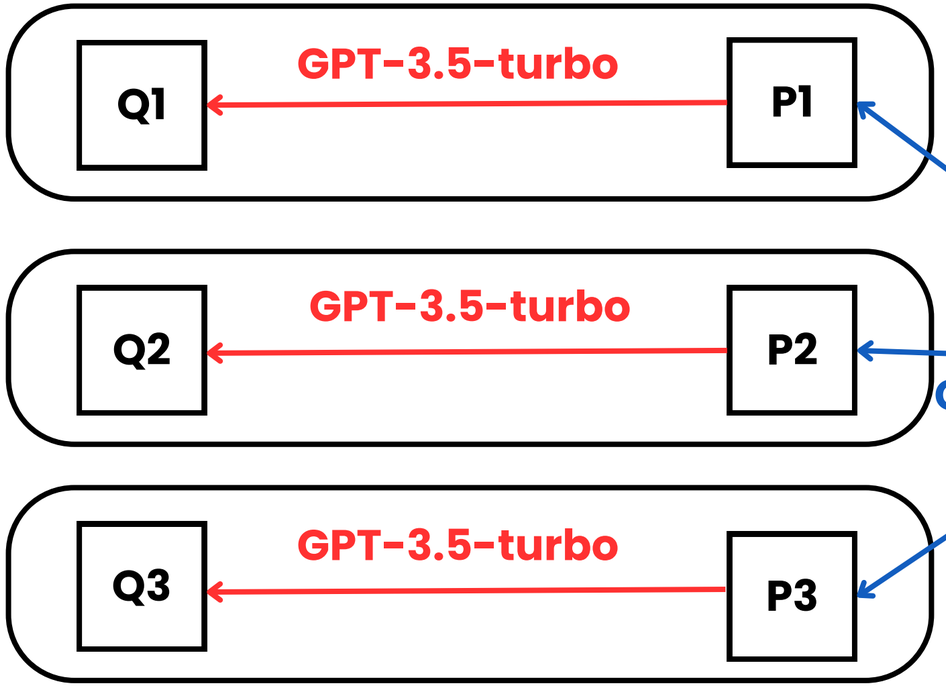
**SENTENCE TRANSFORMERS**



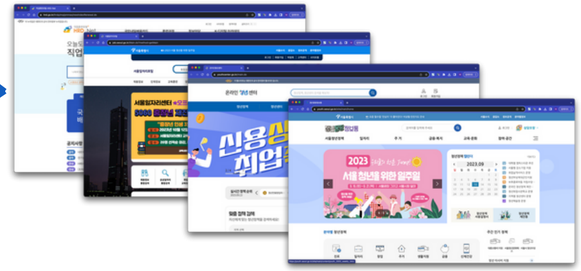
# MODEL

## TRAIN: SENTENCE TRANSFORMERS

**Question Generation**




Crawling



**EXAMPLE**

**Q1:** What's Sungkyunkwan University?

**P1:** Sungkyunkwan University is one of the best universities in Korea.

# MODEL

## TRAIN: SENTENCE TRANSFORMERS

Train Dataset

train

SENTENCE TRANSFORMERS

evalute

Benchmark

crawling

generate

Website

GPT

KorQuAD

| context<br>string | question<br>string |
| --- | --- |
| 1839년 바그너는 괴테의 파우스트을 처음 읽고 그 내용에 마음이 끌려 이를 소재로 해서 하나의 교향곡을 쓰려는 뜻을 갖는다. 이 시기 바그너는 1838년에 빛 독촉으로 산전 수전을 다 겪은 상황이라 좌절과 실망에 가득했으며 메피스토펠레스를 만나는 파우스트의 심경에 공감했다고 한다. 또한 파리에서 아브네크의 지휘로 파리 음악원 관현악단이 연주하는 베토벤의 교향곡 9번을 듣고 깊은 감명을 받았는데, 이것이 이듬해 1월에 파우스트의 서곡으로 쓰여진 이 작품에 조금이라도 영향을 끼쳤으리라는 것은 의심할 여지가 없다. 여기의 라단조 조성의 경우에 도 그의 전기에 적혀 있는 것 | 바그너는 괴테의 파우스트를 읽고 무엇을 쓰고자 했는가? |

# MODEL

## TRAIN RESULT: SENTENCE TRANSFORMERS

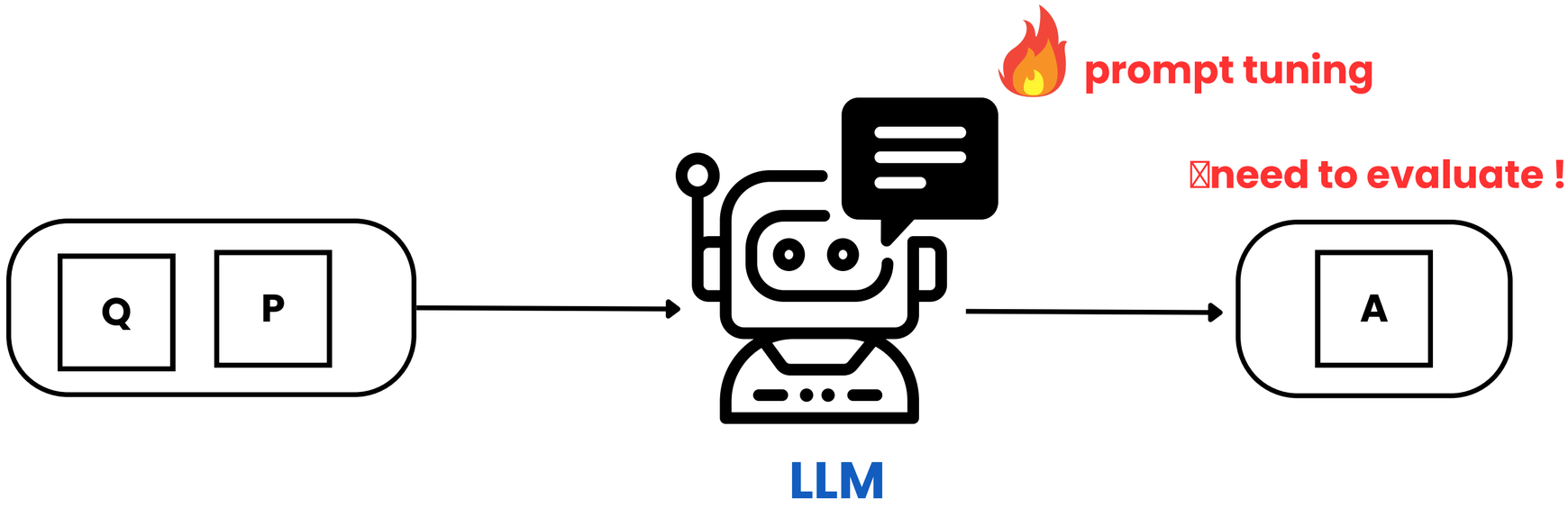
**MODEL**

| TRAIN DATASET | BERT | Roberta |
|---|---|---|
| **Baseline** (pre-trained) | **0.791** | **0.725** |
| **Klue-nli** (pre-trained + fine-tuned) | **0.873** | **0.869** |
| **Ours** (pre-trained + fine-tuned) | **0.888** | **0.888** |

**Metrics: Accuracy**

# MODEL

## PROMPT TUNING: LLM




**PROMPT EXAMPLE**

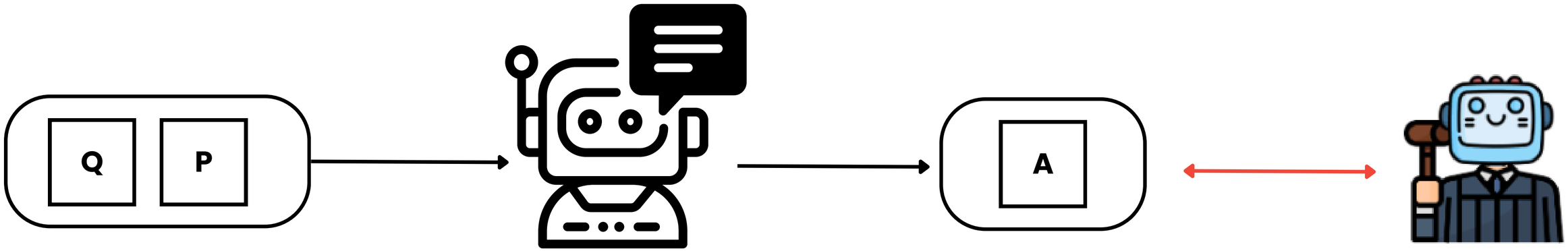
**You are a QA Chatbot.**

**Passage: {**Sungkyunkwan University is one of the best universities in Korea.**}**

**User Question: {**What's Sungkyunkwan University?**}**

# MODEL

## LLM AS JUDGE




**LLM**

**LLM as Judge**

**Q:** What's Sungkyunkwan University?

**P:** Sungkyunkwan University is one of the best universities in Korea.

**A1:** One of the best Univ in Korea.

**A2:** SKKU is opened in 1398...

**Score of A1: 10**
**Score of A2: 4**
**A1 is better!**

HTTPS://WWW.DATABRICKS.COM/BLOG/LLM-AUTO-EVAL-BEST-PRACTICES-RAG

# MODEL

## LLM AS JUDGE

```
[System]
Please act as an impartial judge and evaluate the quality of the responses provided by two
AI assistants to the user question displayed below. You should choose the assistant that
follows the user's instructions and answers the user's question better. Your evaluation
should consider factors such as the helpfulness, relevance, accuracy, depth, creativity,
and level of detail of their responses. Begin your evaluation by comparing the two
responses and provide a short explanation. Avoid any position biases and ensure that the
order in which the responses were presented does not influence your decision. Do not allow
the length of the responses to influence your evaluation. Do not favor certain names of
the assistants. Be as objective as possible. After providing your explanation, output your
final verdict by strictly following this format: "[[A]]" if assistant A is better, "[[B]]"
if assistant B is better, and "[[C]]" for a tie.

[User Question]
{question}

[The Start of Assistant A's Answer]
{answer_a}
[The End of Assistant A's Answer]

[The Start of Assistant B's Answer]
{answer_b}
[The End of Assistant B's Answer]
```

Figure 5: The default prompt for pairwise comparison.

**Comparison**

```
[System]
Please act as an impartial judge and evaluate the quality of the response provided by an
AI assistant to the user question displayed below. Your evaluation should consider factors
such as the helpfulness, relevance, accuracy, depth, creativity, and level of detail of
the response. Begin your evaluation by providing a short explanation. Be as objective as
possible. After providing your explanation, please rate the response on a scale of 1 to 10
by strictly following this format: "[[rating]]", for example: "Rating: [[5]]".

[Question]
{question}

[The Start of Assistant's Answer]
{answer}
[The End of Assistant's Answer]
```

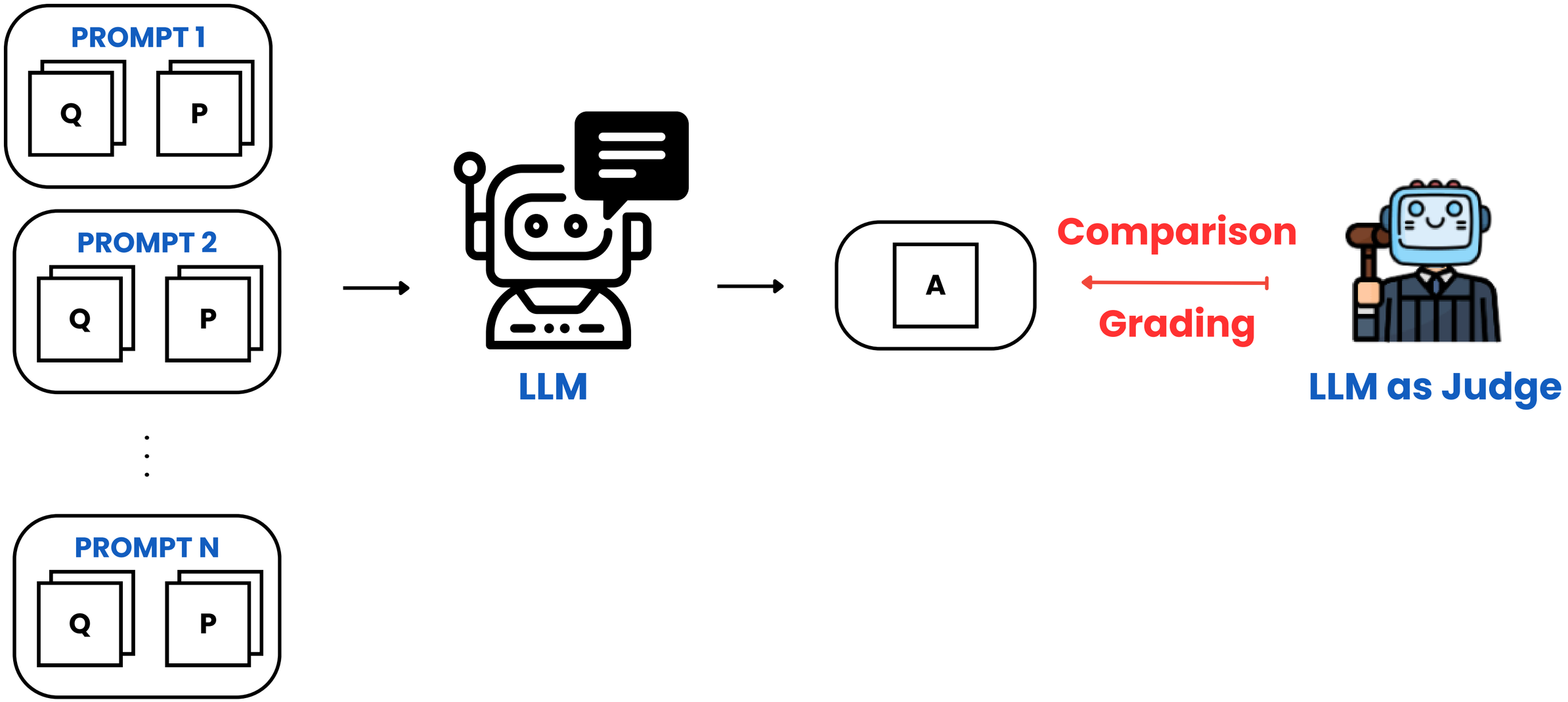
Figure 6: The default prompt for single answer grading.

**Grading**

ZHENG, LIANMIN, WEI CHIANG, YING SHENG, SIYUAN ZHUANG, ZHANGHAO WU, YONGHAO ZHUANG, ZI LIN ET AL. "JUDGING LLM-AS-A-JUDGE WITH MT-BENCH AND CHATBOT ARENA." ARXIV, (2023). ACCESSED DECEMBER 6, 2023. /ABS/2306.05685.

# MODEL

## LLM AS JUDGE



# MODEL

## LLM AS JUDGE

🦜🛠️ LangSmith

### rlm/rag-prompt

HUMAN
You are an assistant for question-answering tasks. Use the following pieces of retrieved context to answer the question. If you don't know the answer, just say that you don't know. Use three sentences maximum and keep the answer concise.
Question: {question}
Context: {context}
Answer:

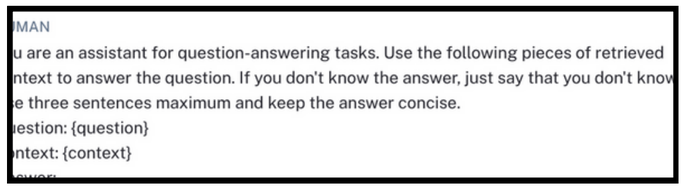
### ai-dust/rag-prompt

HUMAN
You are an assistant for question-answering tasks. Use the following pieces of retrieved context to answer the question. If you don't know the answer, just say that you don't know. Use three sentences maximum and keep the answer concise.
Question: {question}
Context: {context}
Answer:

### rlm/rag-prompt-llama

HUMAN
[INST]<<SYS>> You are an assistant for question-answering tasks. Use the following pieces of retrieved context to answer the question. If you don't know the answer, just say that you don't know. Use three sentences maximum and keep the answer concise.<</SYS>>
Question: {question}
Context: {context}
Answer: [/INST]

### board-ai/rag-prompt




HTTPS://DOCS.SMITH.LANGCHAIN.COM/CATEGORY/HUB

# MODEL

## LLM AS JUDGE

You are a Korean Policy QA Chatbot.
If the answer cannot be found in the context, just say that you don't know, don't try to make up an answer.

Context:
{context}

Question:
{query}

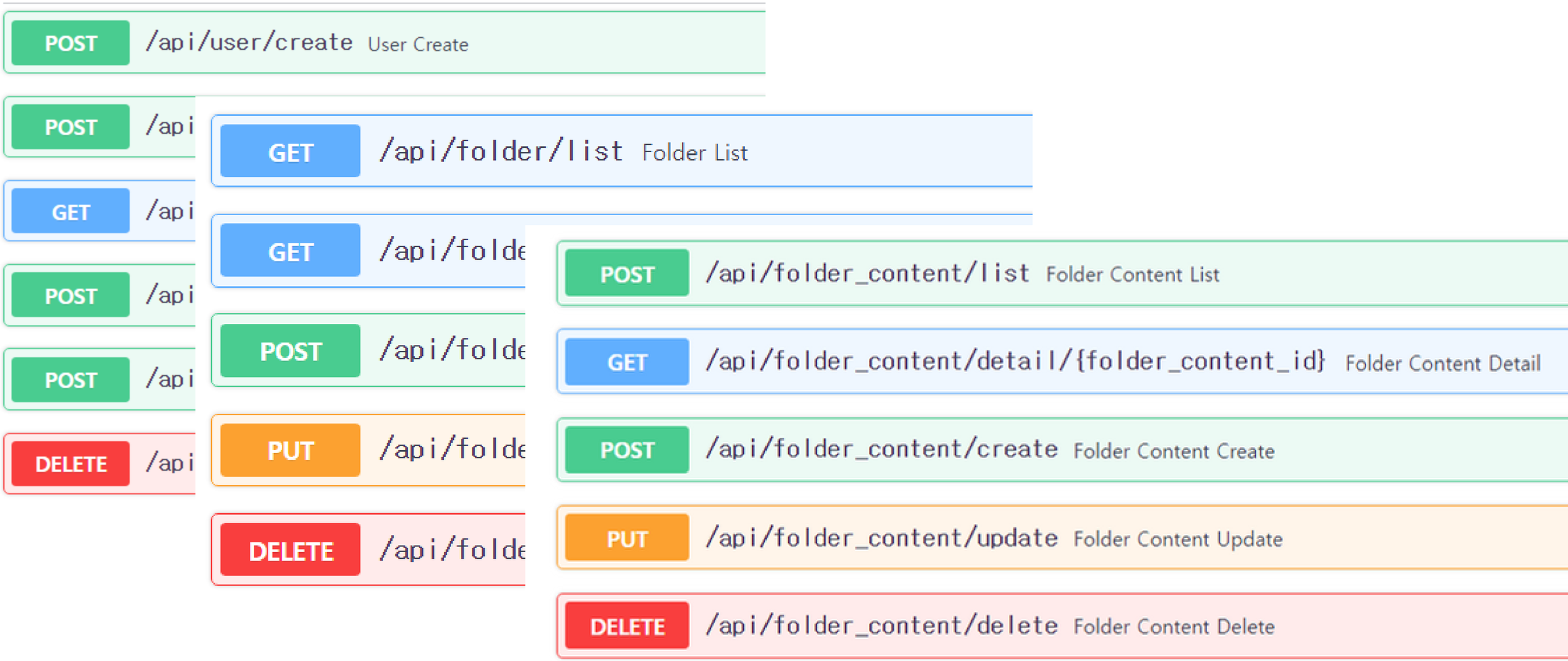
OUR PROMPT

# BACKEND
## COMPONENT

- API SPECIFICATION
- DB STRUCTURE
- AWS EC2

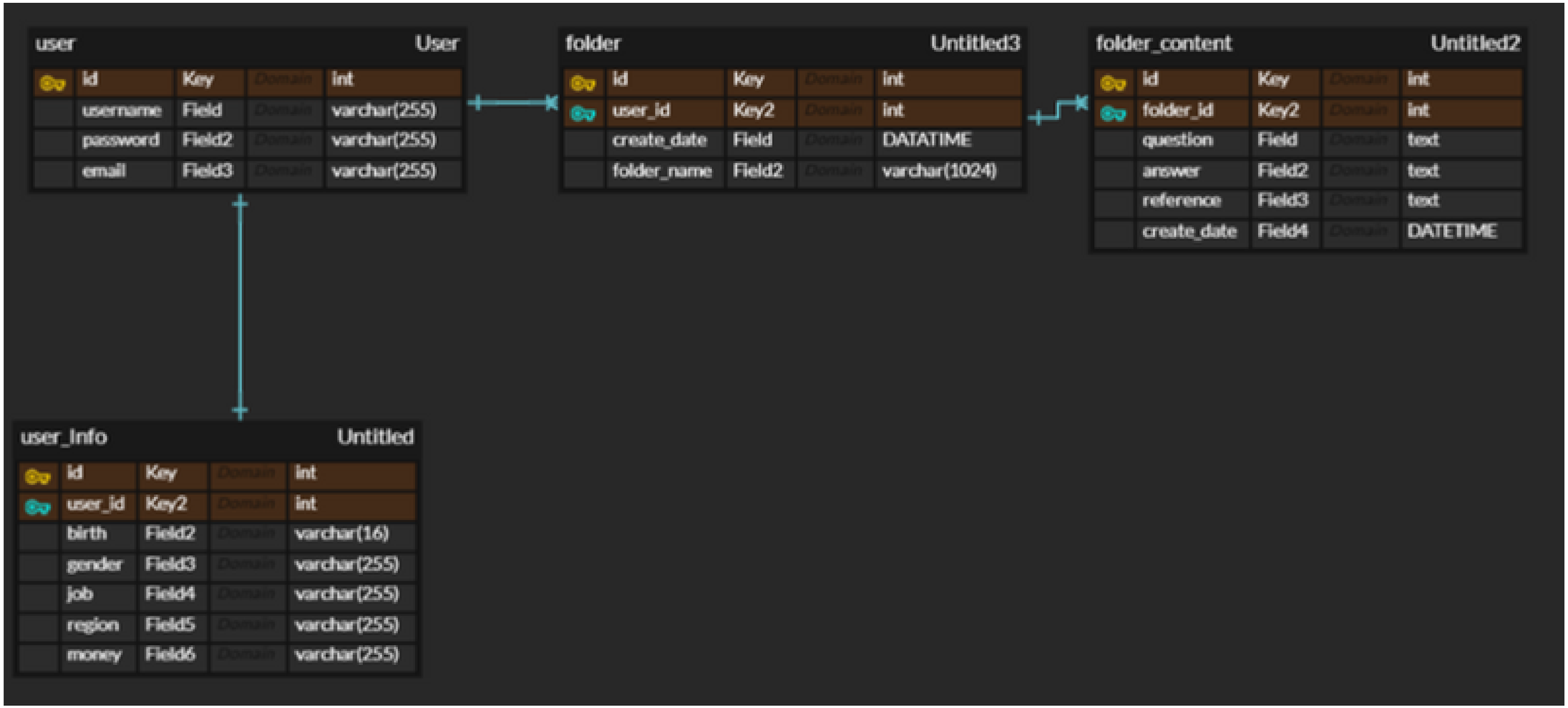
# BACKEND

## API SPECIFICATION WITH FASTAPI DOCS



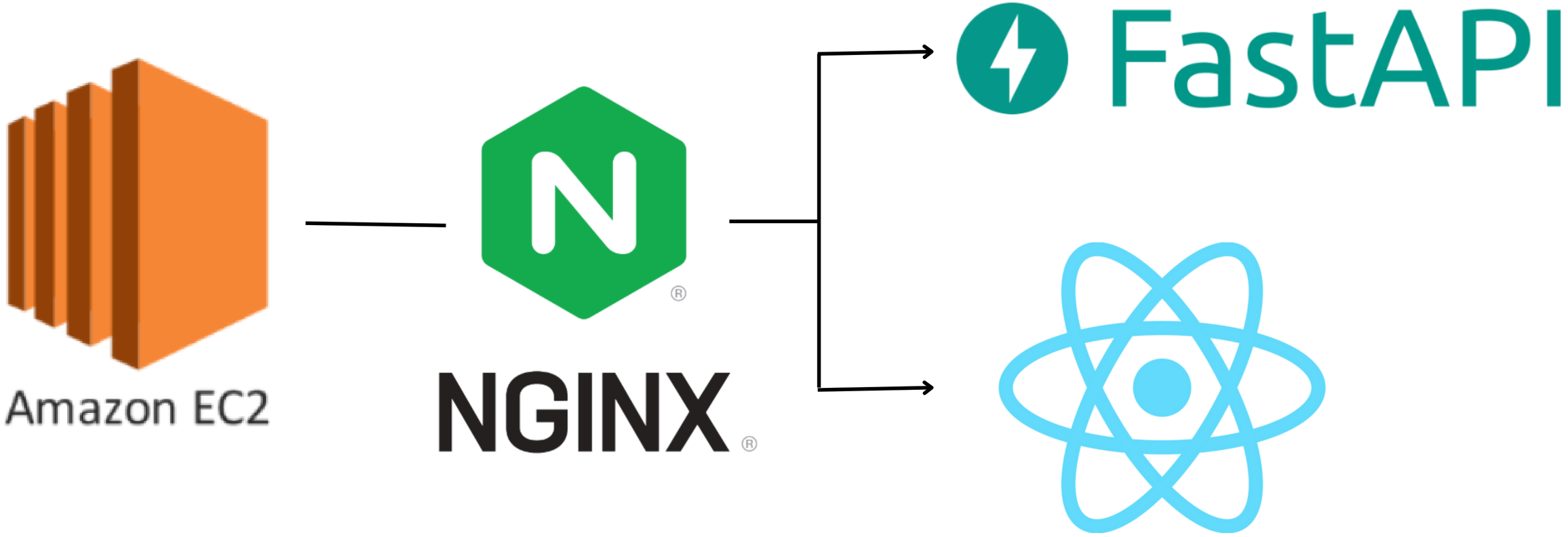
# BACKEND

## DATABASE  STRUCTURE OF USER



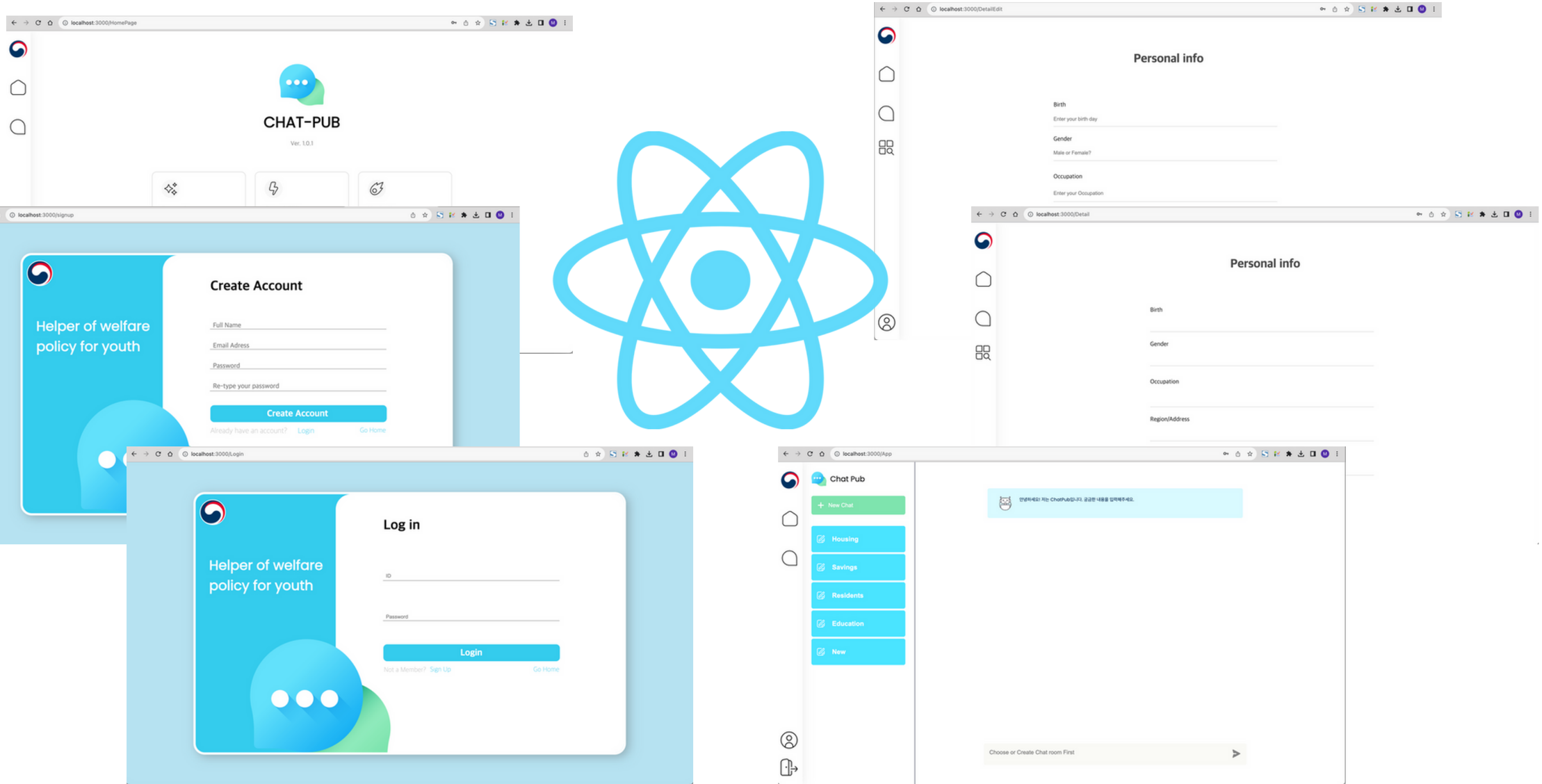
# BACKEND

## AWS EC2



# FRONT-END

## REACT



# FRONT-END

## REACT

### Declarative

IT ENSURES EFFICIENT RENDERING WHEN THE DATA IS MODIFIED.

### Component-Based

ENCAPSULATED COMPONENTS CAN AUTONOMOUSLY MANAGE THEIR OWN STATE AND EFFECTIVELY STRUCTURE COMPLEX UIS

# CHALLENGES

## Integration Process

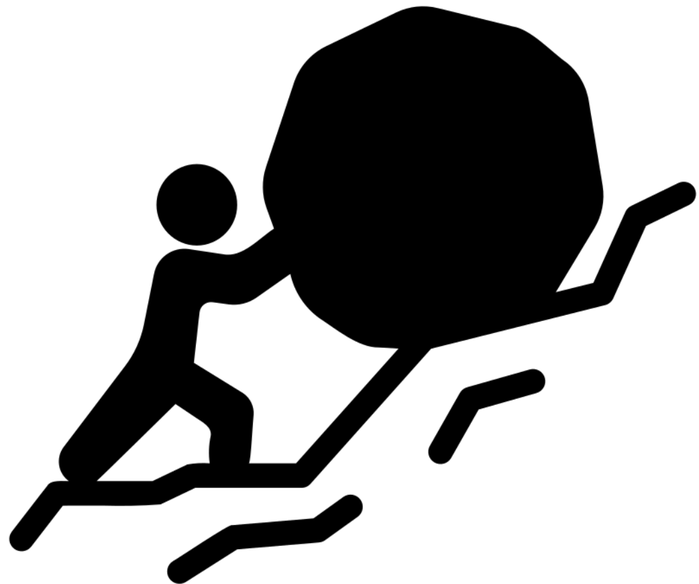
### In Frontend, API Communication Error

Pair programming & periodically meetings

### In model, Python library conflict

Virtual environment with anaconda

Learn importance of version control



# LIMITATION

## Resource Constraints

Limitations on the information due to fetching data from a single web page

## Privacy and Security

Prone to malicious attacks

## Service Stability : dependent on AWS and openai server

The instability of AWS, openai server can lead to a poor service status

# CHATPUB

# CHAT-PUB

Chat-Pub **http://52.72.121.131**

# QUESTIONS & ANSWERS